Automatic Multi-Atlas Segmentation of Myocardium with SVF-Net

Marc-Michel Rohé, Maxime Sermesant and Xavier Pennec

Université Côte d'Azur, Inria, Sophia-Antipolis, France

Abstract. Segmentation of the myocardium is a key step for image guided diagnosis in many cardiac diseases. In this article, we propose an automatic multi-atlas segmentation framework which relies on a very fast registration algorithm trained with convolutional neural networks. The speed of this registration method allows us to use a high number of templates in the multi-atlas segmentation while remaining computationally tractable. The performance of the propose approach is evaluated on a dataset of 100 end-diastolic and end-systolic MRI images of the STACOM 2017 Automated Cardiac Diagnosis Challenge (ACDC).

1 Introduction

Both ventricles play a fundamental role for the circulation of oxygenated blood to the body. To evaluate their functions, clinicians rely on indices that are based on geometrical measurements of regions of the hearts [2, 3]: the blood pool volume, the wall thickness of the myocardium, or the myocardial mass. These indices are usually estimated using a manual segmentation of the contours of the myocardium and the blood pool. However, this task is very time-consuming, requires clinical experience and is prone to large inter-rater variability [9] which will impact the measures derived. For these reasons, there is an important clinical need to define segmentation methods that are fast and fully automatic.

Main challenges to develop such a fully automated segmentation method from medical images are the large variability of the shape of the myocardium, the artifacts and the noise in the images, and the difficulty to chose the most basal slice to segment. In this paper, we propose a method based on multi-atlas segmentation (MAS), an extension of atlas-based methods with multiple templates [7]. With respect to the state-of-the-art MAS methods, our contribution rely on the use of a very fast and robust registration algorithm [6] specifically trained to perform inter-patient heart registration. This registration method leverages recent advances in the field of convolutional neural networks and uses a machine learning approach to the task of registration. The speed of the registration method used paves the way for the use of a high number of atlas which increases the range of anatomy that can be predicted with such a model. One can expect that, for each target geometry we want to segment, at least one or multiple atlases will have a similar shape.

The rest of this article describes our segmentation method step by step. First, as a pre-processing step, we detect the location and the orientation of the

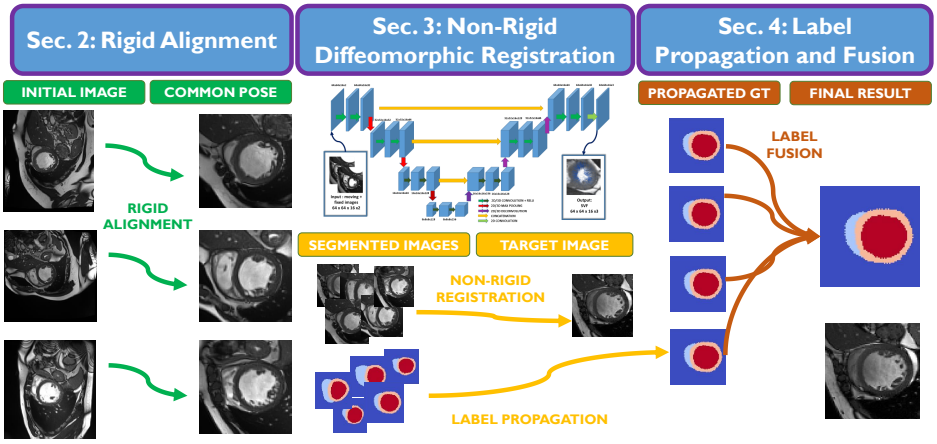

Fig. 1: Overview of our proposed method. Firstly, all the images are aligned in a common pose by computing a $2D$ transformation with 2 landmarks: the center of the LV and th RV. In section 3, we present our non-rigid registration method that is used to register each of the template images with the target image. Finally, in section 4, the labels are propagated and fused to get our final estimation.

heart in the image. We perform this task thanks to a CNN trained to detect two landmarks. One landmark gives the position of the heart and the other one is used to get the orientation. Using these landmarks, the target image is rigidly aligned with respect to the database of templates with ground-truth segmentation. This is a mandatory step before applying non-rigid diffeomorphic registration. In the following section, we present and adapt the SVF-Net registration method [6] to the specific data of the challenge. This registration algorithm performs the non-rigid registration part of our MAS method. Then, we define a method to fuse the label of the estimation from the different templates using specific weights. The pipeline is schematically represented in Fig. 1. Finally, we evaluate our proposed method on the training dataset of the Automatic Cardiac Diagnosis Challenge held in STACOM 2017.

2 Rigid Alignment by Landmarks Detection

For the images of the training database, one can easily define a common pose and alignment by using the barycenter of the LV and the RV computed with the segmentation information. The center of the LV is used to defined a region of interest of the heart while we use the center of the RV to get the axis of the LV to the RV defining orientation of the heart in the X/Y plan. A $2D$ rotation on this plan is applied to all images of the atlases (see Fig. 3) so that RV and LV are aligned.

For the target image, for which we do not have the ground truth segmentation, we need to define a method to detect these 2 landmarks in order to perform

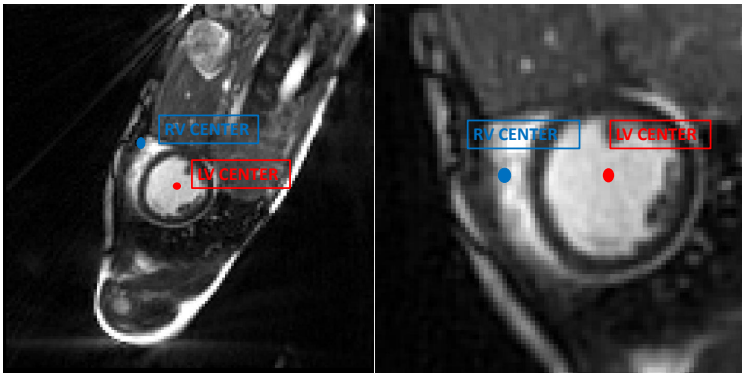

Fig. 2: (Left): Image from a slice of a 3D acquisition. (Right): Same image after pre-processing (cropping the ROI and $2D$ rotation around the Z axis to align with the pose of the atlases). Most of the background has been removed from the image and only the important information remains. The pose of the LV/RV ventricles and the heart position is aligned with the atlas making the registration step easier. To do so, a landmark corresponding to the LV and one corresponding to the RV are detected.

the same pre-processing that was done with the atlases. Inspired by recent works [1], we propose to use heatmaps regression for landmark detection to detect both landmarks (LV and RV centers) using CNNs. In particular, the work of [5] investigates the idea of directly estimating multiple landmark locations from 3D image using a single fully-convolutional CNN, trained in an end-to-end manner to regress heatmaps for landmarks instead of absolute landmark coordinates. This approach has multiple advantages. It is a learning-based method so that we can efficiently leverage our large database of 200 images with ground truth landmarks derived from segmentations. Also, the prediction of heatmaps is an easier task for a CNNs than the prediction of absolutes coordinates of landmarks, as the localization of the responses in the successive layers can be directly used to predict the heatmap.

For all the images of the training set, the heatmap of a landmark with position p is defined on the image grid as:

$$H_p(\mathbf{x}) = \exp\Big(-\frac{\|\mathbf{x}-p\|^2}{\sigma^2}\Big),$$

where σ is the decaying factor of the heatmap. Examples of such heatmap for both the RV center and LV center landmarks are shown in Fig. 3. An CNN U-Net architecture similar to the one presented in [6] is used. The input of the network is the complete image and the output is the predicted heatmap. At test time, the landmark position p is inferred by computing the point that minimizes

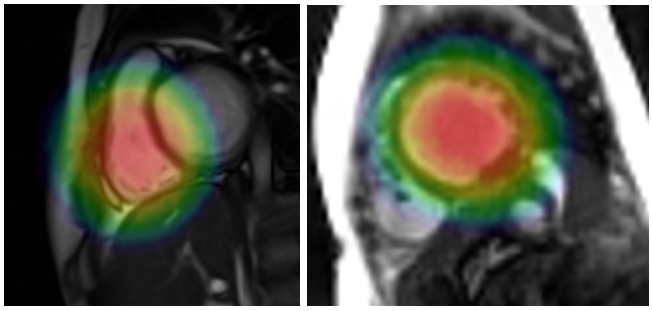

Fig. 3: Heatmaps for both landmarks (RV center: left image, LV center: right image).

the least-square distance to the predicted map H_{pred}:

$$p = \text{argmin} \sum_{\mathbf{x}} \|H_{pred}(\mathbf{x}) - H_p(\mathbf{x})\|^2.$$

3 Non-Rigid Diffeomorphic Registration with SVF-Net

In [6], the authors propose to use Fully-Convolutional Neural Networks (CNN) (illustrated in [6]) to predict directly the deformation from a pair of images. With respect to traditional patch-based approaches, this fully convolutional architecture has the benefit to be faster at test time (registration taking less than 6 sec./30 ms. with CPU/GPU) as the whole image is passed in a single stream to the network instead of passing multiple streams corresponding to the patches of the image in a sliding-window approach. The computational efficiency of this method makes it particularly suitable for our MAS approach paving the way for the use of a large number of templates.

To train this kind of CNN, one needs to compute ground truth registrations from pairs of segmented images. In [6], reference deformations are computed using the result of a registration algorithm previously run on pairs of segmented shape. With respect to the use of the result of the registration on the images, this method to define reference deformations tends to be more robust as the segmentations can be corrected manually. In the dataset provided by the challenge, the segmentations were given in the form of binary masks rather than segmented shapes. Therefore, we adapt the method and perform pair-wise registration of the binary masks using the *LCC-log demons algorithm* [4]. The deformation fields are computed with an iterative optimization run successively on each of the 3 regions of interest.

4 Label Fusion Method

We consider a database of M training images I_j, $j = 1, ..., N$, or atlases for which we have ground truth segmentations (which are images with 3 channels

corresponding to the 3 regions of interest). We perform the registration of each of these images with respect to the target image I with the method described previously. The resulting deformation field is applied to the binary mask of the atlas M_j and we need to define a method to combine these estimations $\hat{M_j}$ to get $\hat{M_j}$: the estimation of the segmentation of the target image.

A straight-forward method to combine the estimations is by majority voting. In this work, we chose to use varying decision weights in order to combine these estimations using a local assessment of the registration success. Therefore, these weights will have a higher value for registrations in which we have a higher confidence. This confidence is evaluated at each point of the image using 3 different metrics. The first one is the *Local Correlation Coefficient* (LCC) [4] which locally estimates the similarity between the voxel intensities of the images. The second metric is the square norm of the displacement of the transformation because we consider that we have stronger confidence in small transformations (corresponding to similar images) than in large transformations. Finally, the last metric corresponds to the Jacobian of the deformation field, meaning we give more confidence to smooth displacement fields over non-regular ones.

To get the function $d_j(x)$ representing the local assessment of the registration success (between the target image and atlas j), these 3 metrics are combined linearly. The coefficients are learned on the training set as to minimize the square difference of the distance versus the ground truth labeling error. Then, we define the local weight of each point of each atlas as a function of the local distance with a kernel σ_{metric}. Furthermore, we also smooth the weights spatially with a kernel $\sigma_{spatial}$ in order to ensure spatial consistency of the estimations. Finally, the weights are normalized in order to sum to 1 :

$$\tilde{\omega}_j(x) = G_{\sigma_{spatial}} \star \exp\left(-d_j(x)/\sigma^2_{metric}\right),$$

$$\omega_j(x) = \frac{\tilde{\omega}_j(x)}{\sum_k \tilde{\omega}_k(x)}.$$

The kernel σ_{metric} corresponds to the confidence on our estimation of the local distance $d(\hat{p}^j_k)$. Large values for the kernel corresponds to small confidence. At the limit when σ_{metric} becomes large enough, all the weights become equal and we get the simple method of averaging the labels and the local distance does not have any impact. The spatial kernel $\sigma_{spatial}$ is to ensure spatial consistency of the resulting segmentation. Large values of $\sigma_{spatial}$ will make the weights more global whereas small values will make them more local.

5 Results and Discussion

Our method is applied to the database of the STACOM 2017 Automated Cardiac Diagnosis Challenge (ACDC). This challenge provides the community with a comprehensive set of $3D$ cine-MRI images (100 patients divided in 5 groups: 4 pathological plus 1 healthy control groups) acquired at the University Hospital of Dijon. For each of these patients, manual expert segmentation was performed for

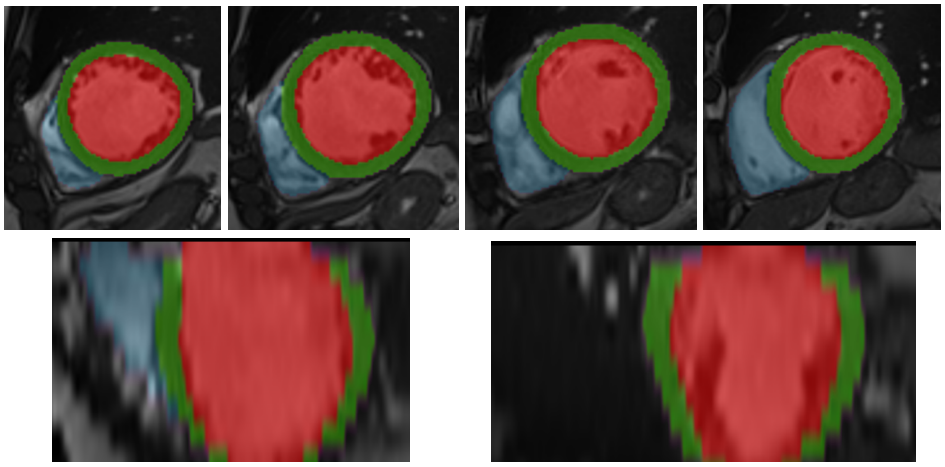

Fig. 4: Example of a segmentation using proposed method. (Top): short axis view with 4 slices. (Bottom): 4CH and 2CH axis views.

the end-diastolic (ED) and end-systolic (ES) frame to trace the LV endocardial and epicardial borders giving 3 regions of interest: RV cavity, LV cavity and left ventricle myocardium. We use 80 patients to train the CNN networks and 20 patients as testing test to evaluate the accuracy of the segmentation.

Training. Reference deformations are computed for each of the possible combinaison of pairs of our dataset of 80 ED and 80 ES images for a total of $160^2 = 40,000$ reference deformations which took 2 minutes per pair on a single core CPU (a cluster of CPU was used). Because our method already gives us a large database of ground truth data, we only use small translations in the X and Y axis for data augmentation (this also improves the robustness of the learned network over slight rigid misalignment of both images). For the loss function, we used the sum of squared difference beetween the predicted SVF parametrization and the ground truth. We implement the network using Tensorflow[1] and we train it on a NVIDIA TitanX GPU with $100,000$ iterations using the ADAM solver which took approximately 24 hours. The CNN to detect the position and the orientation of the heart is trained similarly using these 160 images. The coefficients of the function $d_j(x)$ are then learned using leave-one-out for each of the testing images (19 testing images are used to estimate the optimal coefficients that is applied to the other image). Finally, σ_{metric} and $\sigma_{spatial}$ are estimated with a trial and error approach to balance accuracy and smoothness of the result.

Testing. We evaluate the method on the 20 ED and ES testing images. For each of these images, we perform the registration using SVF-net with respect to

[1] www.tensorflow.org

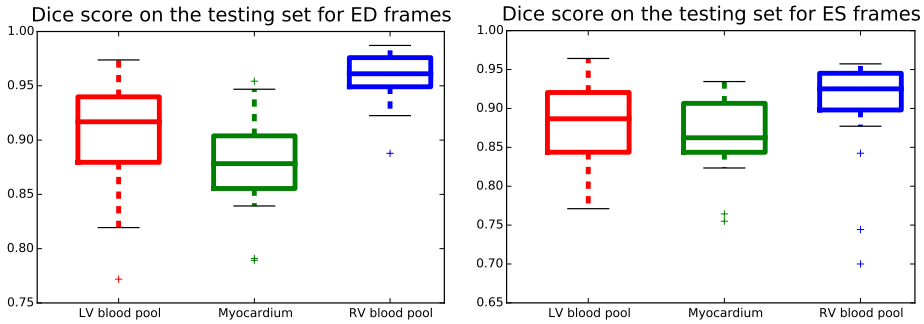

Fig. 5: Results on the 20 patients used for testing. (Left): dice scores for the ED frames. (Right): dice scores for the ES frames. The three different regions of interests are shown.

the 80 training images and fuse the warped labels using the method described in section 4 to get the final estimation of the label corresponding to the regions of interest. An example of the segmentation with our method can be seen in Fig. 4. One can see that our method produces a segmentation that is smooth and spatially consistent in the Z axis. When compared qualitatively to ground truth segmentations, most of the differences were seen at the base, where our method did not always segment the same basal slice as the ground truth. Additional work could be done to our method so that we come up with a more consistent evaluation of the first slice to segment. Finally, we have evaluated quantitatively the results using dice scores in Fig. 5. As expected, dice scores for LV blood pool (median of 0.97/0.93 for the ED/ES frames) tend to be higher than for the two other regions of interest with myocardium at 0.87/0.87 and RV blood pool at 0.92/0.89 for ED/ES. These results are promising and need to be confirmed and compared to other state-of-the-art methods on the final testing set of the challenge.

Extension to classification. The function d_j defined in section 4 represents the distance between pairs of myocardium shapes. This distance can be used, together with more advanced statistics of deformation fields, to perform classification of the target patient. For example by looking for the closest patients with respect to this distance or by running classical machine learning algorithms on the deformation fields corresponding to the pairwise registrations.

6 Conclusion

In this article, we present a method for the segmentation of the myocardium using multi-atlas segmentation. The method we present has several important qualities. It is completely automatic and does not even require the location of the

heart as user input, thanks to the landmarks detection network. With respect to traditional multi-atlas segmentation algorithm, the speed of the registration method allows us to use a large database of atlases while keeping the method computationally tractable. To combine the different segmentation into the final result, we define local weights that are a priori learned on a training sample. These weights are based on an estimation of the confidence of the evaluation of a specific point by each atlas. These weights and the deformation fields of the registration result could be used to perform classification of patients with respect to the 5 classes provided by the challenge dataset. The method is evaluated on the training set of the ACDC challenge and is ready to be applied to the final testing dataset.

References

1. Bulat, A., Tzimiropoulos, G.: Convolutional aggregation of local evidence for large pose face alignment. British Machine Vision Conference (2016)
2. Kilner, P.J., Geva, T., Kaemmerer, H., Trindade, P.T., Schwitter, J., Webb, G.D.: Recommendations for cardiovascular magnetic resonance in adults with congenital heart disease from the respective working groups of the european society of cardiology. European heart journal p. ehp586 (2010)
3. Kramer, C.M., Barkhausen, J., Flamm, S.D., Kim, R.J., Nagel, E.: Standardized cardiovascular magnetic resonance (cmr) protocols 2013 update. Journal of Cardiovascular Magnetic Resonance 15(1), 91 (2013)
4. Lorenzi, M., Ayache, N., Frisoni, G.B., Pennec, X.: Lcc-demons: a robust and accurate symmetric diffeomorphic registration algorithm. NeuroImage 81, 470–483 (2013)
5. Payer, C., Štern, D., Bischof, H., Urschler, M.: Regressing heatmaps for multiple landmark localization using cnns. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 230–238. Springer (2016)
6. Rohé, M.M., Sermesant, M., Pennec, X.: SVF-Net: Learning Deformable Image Registration Using Shape Matching. In: MICCAI 2017 - the 20th International Conference on Medical Image Computing and Computer Assisted Intervention. MICCAI 2017, Lecture Notes in Computer Science, Quebec, Canada
7. Rohlfing, T., Brandt, R., Menzel, R., Maurer, C.R.: Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. NeuroImage 21(4), 1428–1442 (2004)
8. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234–241. Springer (2015)
9. Suinesiaputra, A., Bluemke, D.A., Cowan, B.R., Friedrich, M.G., Kramer, C.M., Kwong, R., Plein, S., Schulz-Menger, J., Westenberg, J.J., Young, A.A., et al.: Quantification of lv function and mass by cardiovascular magnetic resonance: multi-center variability and consensus contours. Journal of Cardiovascular Magnetic Resonance 17(1), 63 (2015)